The Efficacy of Non-Probability Online Samples

Alan Roshwalb, Zachary Lewis and Robert Petrin

Ipsos Public Affairs, 202 K Street NW, Washington, D.C. 20006

Abstract

The use of non-probability online opt-in samples is very common in commercial marketing and survey research practice. The use of non-probability online samples in experimental studies is well understood and without concern. When their use is for population estimation, it is here that classical survey statistical theory has concerns in their use. Given the cost, speed and flexibility of online surveys, non-probability samples are being used more and more for the measurement and tracking of attitudes and behaviors. In this study, we examine two non-probability online sample surveys for the purpose of exploring strategies to adjust and calibrate non-probability samples. Post-survey weighting is examined including standard raking ratio, propensity, and Generalized Regression weighting. The different auxiliary variables used in the adjustments include standard demographics, technology adoption, and components of an attitudinal-behavioral consistency strategic model. The attitudinal-behavioral consistency model motivates the discussion and provides guidance for measures to include in adjustments. It is under this framework that we examine the efficacy of non-probability online samples.

Key words: Non-probability samples, calibration methods, attitude-behavior consistency model.

1. Introduction

For the last 15 years, survey research has seen the use of online samples in marketing, survey and public polling research expand greatly. At first, the use of online surveys was mostly in controlled experiments often found in product, message and concept testing in marketing research. Today, online surveys are found in public opinion polls tracking trends in attitudes towards social issues, politics, elections, and, of course, goods and services. Recently, U.S. Government agencies began using online samples to track some attitudes and behaviors.

Online surveys most often use samples drawn from panels of individuals or households that have either volunteered themselves or have been recruited to participate in surveys. In most cases, the recruitment process is a non-probabilistic online intercept through advertisements or pop-up request. A panel then serves as a sampling frame. Since the sample frames are constructed using non-probability methods, the sample selected from these frames are non-probability samples. In addition to the panels, some companies use blended samples with part of the sample coming from a panel and another part coming from real-time recruitment off the Internet, often called river sampling.

Consumer panels for marketing research have been around since the early 1950's with the original ones created for mail survey research (Sudman and Wansink, 2002). The efficacy and usefulness of the mail panels has been noted several times (Groeneman, 1991; Putnam, 2000). More recently, these panels and now opt-in web panels recruit new panel members using online promotions and invitations. Once they have agreed to become members, recruits complete questionnaires and agree to participate in surveys through email invitations (Couper and Bosnjak, 2010).

Over time, concerns and shortcomings with on-line panels have come to light. The concerns include the properties of the panels of these volunteers from skewed distribution of panel members, the limited size of the panels, and achieved response rates. A growing body of work has been focused on methods for correcting the potential biases in these panels as a way to improve their accuracy and usefulness.

The greatest concern among practitioners and users of the survey data from online studies is the efficacy of these panels and properties of samples in terms of representativeness and bias. In 2010, AAPOR published a task force report on Online Panels that noted that inferences from non-probability online panels tended to be less accurate than those from probability samples (Baker et al. 2010).

In this paper, we look at the online panels properties and examine the methods to adjust data from online panels to manage potential biases. We explore results for two online panel studies. One was conducted solely for the purpose of this research, and the other was run to see if one could migrate a study from an RDD telephone survey to an online panel study. Some of the results are promising, but we still think the use of non-probability samples is still a cautionary tale.

2. Methodological Backdrop

Efficacy, as used here, relates to the representativeness and the ability to cope with potential biases in a sample to study a population. Without question, probability sampling is the most appropriate approach. A good probability sample creates a cross-section of the population, and classical statistical sampling theory provides the means for describing and assessing design-based unbiasedness and design-based variance. Under ideal circumstances, the coverage of a sample frame has nearly 100% coverage of the population. As a result, design-based unbiasedness of sample statistics has expectations equal or just negligibly different from population parameters under random sampling captured in the probabilities of inclusion.

A common criticism of non-probability samples is that it is not possible to determine probabilities of inclusion, because panelists were not recruited at random. Studies using non-probability sampling cannot rely on classical statistical sampling theory for establishing representativeness and provide estimates of margins of error.

Some researchers rely on Bayesian statistics to provide a framework to discuss representativeness and variance measures in non-probability sampling. Uncertainty of an estimate is based on what one knows about the population relying either on super-population models or Bayes theorem to generalize from the sample to the population[1]. Bayes statistics allows sample designs to be conditionally ignorable such that a sample can come from a probability or a non-probability sample as long as the parameters under likelihood are consistent with the parameters of the prior. Therefore, unbalanced samples from non-response, non-coverage or other nonrandom nature of designs can be conditionally ignorable if the samples can be balanced. Elements of this discussion have been used in promoting the use of Bayesian credible Intervals as measures of the degree of certainty of parameters from non-probability samples (Roshwalb, El-Dash, and Young, 2012; AAPOR Statement on Credibility Intervals, 2012).

2.1 Attitudinal models

The most common approaches in the adjustment of non-probability samples focuses on adjusting a sample through post-field methods to account for differences between the general population and those people that were accessed by the study. The earliest attempts “fixed” the sample in terms of demographics. Additional components such as technology adoption and technology usage were included. Most of these approaches did not account for differences in the way people think, behave and create attitudes.

Conceptually, one must incorporate how individuals’ attitudes and their behavior evolves relative to basic beliefs, actions and attitudes into an adjustment structure. We use an attitude-behavior consistency (ABC) model to provide structure for discussion and choosing measures to be included.

The ABC model was first discussed in the 1930’s as “the affect for or against a psychological object” model (Thurstone, 1931). The ABC is composed of three components (Rosenberg and Hovland, 1960):

- Affective – involves a person’s feelings about an attitude object,
- Conative – involves the way an attitude influences how a person acts or behaves, and

[1] For a discussion of probabilistic sampling from a Bayesian perspective see Kadane and Seidenfeld.(1990).

- Cognitive – involves the way a person's belief or knowledge about an attitude object.

The model requires that the people's behavior must be rationally consistent with their attitudes. This linkage was discussed in the context of marketing research by Bagozzi, et al (1979). We use this in the context of survey research and subpopulations. By adjusting by the ABC model, differences in the ABC model components between those that willing join a panels versus the general population should be able to balance samples in terms of the measurement of attitudes.

The implications of the ABC are even more important when considering that Greenwald (1968) noted that each component has distinct learning processes and influences. In other words, if specific subpopulations are subject to different experiences, means of acquiring information, and acting, then their attitudes and subsequent behaviors can be distinctly different than subpopulation. As a result, the online population may be different than the non-online population because of their means of learning and actions, and the population that choose to volunteer to respond to surveys online versus those that do not as well.

2.2 Correction approaches

Methods used to realign and correct the sample once the data have collected are raking-ratio, propensity, and Generalized Regression weighting. The relative merits of each will be examined, but we will explore the value of including demographics, technology use and adoption and measures related to the ABC model.

For online sampling, we assume pseudo-probability sampling structure (AAPOR 2013). Sample weights are assigned to correct for imbalances in the sample relative to population targets such as age, gender, race, etc.

The underlying model in all of the approaches is based on the following. Estimates from the sample are obtained using the standard formula for a weighted mean. It is

$$\hat{T} = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i}. \tag{1}$$

W_i is the sample weight. Without any adjustment, it is the projection rate for the sample. The process evolves when weighting adjustments are included. The weights usually become a function of the projection rate and the adjustments. The formula for the weight is written as:

$$W_i = u(Z_i)/p_i(s). \tag{2}$$

Here, $1/p_i(s)$ are the projection rates for the sample, and $u(Z_i)$ are the weighting adjustment factors for a set weighting variables Z.

2.2.1 Raking-Ratio Weighting

Early on in survey research, estimates of variables from well-constructed survey designs with properly constructed weights were not close to known population statistics. These differences could be due to the randomness of the sample, nonresponse, or under-coverage. Raking-Ratio weighting, or sometimes called rim weighting, is a favored approach. It is easy to implement, and it can reduce biases from nonresponse and non-coverage in sample surveys. Raking ratio weighting adjusts sampling weights so that the weighted sample totals for key categorical variables match the totals for the population (Deming 1943; Kalton 1983). Raking uses iterative proportionate fitting. Each iteration's weights are adjusted sequentially to known marginal totals of the target variables until there is little change in the weights from iteration to iteration. Raking ratio algorithms are readily available in R, SAS and for other data analysis systems. Raking-ratio weighting can implicitly estimate the weights W_i and its components, $u(Z_i)$.

2.2.2 Propensity Weighting

Propensity weighting relies on the ability to determine whether there are differences in attributes of respondents in a non-probability sample from those in a probability sample. Capture these in a propensity model to help balance the non-probability sample. Propensity weighting often uses logistic regression to estimate the probabilities that different respondents will come from a probability sample versus a non-probability sample. Noted properties of propensity weighting are:

- A logistic regression model estimates the probability of response from probability versus non-probability sample, so the information included in the model needs to be available for the two types of samples.
- Categorical variables can be included in the model, but the adjustments may be similar.
- Continuous or scalar variables can be included in the logistic model.

The underlying logistic model takes on the form of

$$\ln\left(\frac{v(Z_i)}{1-v(Z_i)}\right) = \beta_o + \sum_{k=1}^{K} \beta_k Z_{k,i}, \text{ or}$$

$$v(Z_i) = \frac{e^{\beta_o + \sum_{k=1}^{K} \beta_k Z_{k,i}}}{1 + e^{\beta_o + \sum_{k=1}^{K} \beta_k Z_{k,i}}}.$$

The weight for propensity weighting is $W_i = u(Z_i)/p_i(s)$ where $u(Z_i) = 1/v(Z_i)$.

2.2.3 Generalized Regression Weighting

Generalized Regression (GREG) weighting was first proposed in 1992, and it is a systematic approach to using auxiliary information in the adjustment of the weights (Deville and Sarndal, 1992). As with raking, GREG weighting forces the weighted sum of each of a set of variables to equal specified targets. GREG weighting uses an underlying linear prediction model, and it implicitly estimates the inverse of its probability of response. In many ways, GREG is similar to raking-ratio. It is an iterative process, and it can fit the data to categorical or ordinal weighting variables. Where it differs is that it uses a different iterative fitting process, and this process allows for the inclusion of continuous weighting variables. GREG weighting is being used in a variety of government studies from labor force, forestry, and agricultural studies among others. The underlying algorithms are more complicated than raking or propensity weighting, and the SAS CALMAR macro is a comprehensive algorithm (CALibration on MARgins or CALMAR) based on Deville and Sarndal (1992).

3. Results

Two studies are included in this research. Two separate online studies were conducted using questions from the National Survey on Drug Use and Health (NSDUH)[2] and Sallie Mae's How American Pays for College Study[3]. We chose two survey questions from each study to compare their results from an online sample versus those of a probability sample. The surveys and their measures are:

1) National Survey on Drug Use and Health (NSDUH) study Alcohol Risk measures are:
    a. Risk of harming oneself with the heavy consumption of alcoholic beverages, and
    b. Average age of some having their first alcoholic drink.

NSDUH is a national probability sample in-home survey where the respondent answers most questions in private and enters their responses directly into a computer. Here the mode of NSDUH is almost the same as an online survey. The latest NSDUH results are used as the gold standard target results for this data.

2) Sallie Mae's How American Pays for College Study Attitude towards College measures are:
    a. College is an investment in the future, and
    b. A college degree is more important now than it used to be.

---

[2] https://nsduhweb.rti.org/respweb/project_description.html

[3] https://www.salliemae.com/plan-for-college/how-america-pays-for-college/

The Sallie Mae's How America Pays for College Study is a national telephone study conducted once a year using RDD among college students and parents of college students. Here the mode of the national probability survey is interviewer administered by phone versus self-administered online. Here the 2015 Sallie Mae's How America Pays for College study results were used as the gold standard target results for this data.

In addition to these questions, we used questions from the Health Interview National Trends Survey (HINTS) and the General Social Survey (GSS)[4] to serve as measures for the ABC model. These questions were chosen based partly on their relevance to the ABC model, availability of either published values or data sets, and their survey administration mode, mail or in-person self-administered CAPI. The questions are:

Table 1: ABC Model Questions and Survey Source

ABC Component	Question	Survey Source
Attitudinal	Spending too much money on it, too little money, or about the right amount on improving the environment?	GSS
	Spending too much money on it, too little money, or about the right amount on improving education?	GSS
	Spending too much money on it, too little money, or about the right amount on dealing with drug rehabilitation?	GSS
Cognitive (Belief)	People should be willing to help others who are less fortunate	GSS
	Those in need have to learn to take care of themselves and not depend on others	GSS
Conative (Behavioral)	Given food or money to a homeless person?	GSS
	Done volunteer work for a charity?	GSS
	Given money to a charity?	GSS
	Given directions to a stranger?	GSS
Technology and Technology Adoption	When you use the Internet, how do you access it? A regular dial-up telephone line Broadband such as DSL, cable or FiOS A cellular network (i.e., phone, 3G/4G A wireless network (Wi-Fi)	HINTS
	About how many minutes or hours per week do you spend sending and answering electronic mail or e-mail?	GSS

The Alcohol Risk study consisted of 2010 respondents surveyed in May, 2016, and the Attitudes towards College Study consisted of 643 respondents surveyed in June, 2015. These studies used blended samples with panelists selected directly from Ipsos's i-Say online panel, panel members from other panel partner companies, along with people invited in real-time from social web sites.

The next table provides results for each survey question. The Target % represents the calculations from the actual survey, and the Survey % is the results calculated from the online survey using basic demographic weights.

The Online results are not too different for most of the variables. There are some differences especially for the Technology Adoption questions.

[4] The General Social Survey (GSS) is a project of the independent research organization NORC at the University of Chicago, with principal funding from the National Science Foundation.

Table 2: Weighting Variables Target and Actual Survey Results

		Alcohol Risk		Attitude towards College	
Question	Response Categories	Target %	Survey %	Target %	Survey %
NotEnvir	Improving the environment				
1	Too Little	58.7	51.1	19.6	15.0
2	About Right	28.4	27.4	39.5	53.7
3	Too much	8.3	13.2	33.8	31.3
8	DK	4.6	8.3	0.96	1.00
NatEduc	Improving education				
1	Too Little	61.7	62.4	61.4	50.5
2	About Right	29.0	21.7	21.7	38.7
3	Too much	6.8	9.7	10.0	10.6
8	DK	2.5	6.2	6.9	1.0
NatDrug					
1	Too Little	59.4	46.9		
2	About Right	27.9	27.8		
3	Too much	7.8	12.1		
8	DK	4.9	13.3		
GiveHomeless	Give food or money to homeless person				
1	More_than_once_a_week	2.5	5.7	2.5	2.4
2	Once_a_week	4.0	5.4	2.6	6.1
3	Once_a_month	11.7	11.7	16.9	10.2
4	At_Least 2 or 3_times_in_the_past_year	31.6	20.5	24.2	25.6
5	Once_in_past_year	14.6	17.0	16.5	10.2
6	Not_at_all_in_the_past_year	35.5	39.7	37.4	45.5
Directns	Give directions to stranger				
1	More_than_once_a_week	5.3	6.9	9.2	6.2
2	Once_a_week	6.0	7.7	8.1	8.0
3	Once_a_month	17.8	16.1	19.7	15.7
4	At_Least 2or3_times_in_the_past_year	47.0	30.2	33.5	36.7
5	Once_in_past_year	12.7	16.6	10.9	16.1
6	Not_at_all_in_the_past_year	11.2	22.5	18.6	17.3
VolCharty	Done volunteer work				
1	More_than_once_a_week	4.5	6.8	5.5	4.5
2	Once_a_week	3.6	7.8	12.3	10.2
3	Once_a_month	9.2	11.7	15.7	14.9
4	At_Least 2or3_times_in_the_past_year	16.9	14.6	23.0	17.2
5	Once_in_past_year	12.7	13.0	11.1	9.0
6	Not_at_all_in_the_past_year	53.1	46.1	32.3	44.1
GiveChrty	Give money to charity				

Table 2: Weighting Variables Target and Actual Survey Results

		Alcohol Risk		Attitude towards College	
Question	**Response Categories**	**Target %**	**Survey %**	**Target %**	**Survey %**
1	More_than_once_a_week	2.7	5.9	1.1	4.2
2	Once_a_week	8.5	10.0	16.7	15.6
3	Once_a_month	19.9	20.4	29	16.4
4	At_Least 2or3_times_in_the_past_year	33.2	27.3	33.8	22.9
5	Once_in_past_year	14.8	15.9	5.9	16.7
6	Not_at_all_in_the_past_year	21.0	20.6	13.6	24.0
Othshelp	People willing to help others				
5	StronglyAgree	43.4	38.0	51.1	41.2
4	Agree	46.3	44.6	44.2	46.7
3	NeitherAgreeNorDisagree	8.7	14.5	2.1	11.5
2	Disagree	0.9	1.3	1.1	.4
1	StronglyDisagree	0.8	1.6	1.6	.3
Careself	Need to take care of oneself				
5	StronglyAgree	10.5	15.6	19.3	13.1
4	Agree	40.3	31.4	33.1	39.0
3	NeitherAgreeNorDisagree	26.0	33.6	19.5	31.5
2	Disagree	19.4	14.6	19.9	12.6
1	StronglyDisagree	3.9	4.8	8.3	3.9
B2aDialup	Internet through dialup				
1	Yes	3.4	13.8		
2	No	96.6	86.2		
B2bDSL	Internet through broadband/DSL				
1	Yes	63.9	71.4		
2	No	36.1	28.6		
B2cCellular	Internet through Cellular				
1	Yes	59.8	64.2		
2	No	40.2	35.8		
B2dwireless	Internet through wireless				
1	Yes	78.7	81.0		
2	No	21.3	19.0		
FIINOPhones	How you and family receive calls				
1	AllCallsOnell	34.4	53.4		
2	SomeCellAndSomeLL	40.4	32.8		
3	FewCallsOnCell	25.2	13.8		

The next table shows the improvements from adjustments when using the ABC model components. The unweighted results for Alcohol Risk questions are 2% different than the reported Target %'s and 1 year different in the average drinking age. Weighting by demos alone pushed the non-probability results further away from the target numbers. Raking using GSS questions along with the demographic questions and technology adoption questions pushed the results towards the target results.

The results for the Attitudes towards Colleges show that the Target Values measured using a RDD telephone study had far fewer responses in the middle of the scale versus the self-administered online survey. Survey research often sees this pattern when observing mode response effects. The weighting with GSS improves the distribution to some degree. Raking, once again, had the best performance of the adjustment methods, but the results are still different enough to be concerned about the comparability.

Table 3: Results from Adjustment

Outcome Questions		**Target**	**Un-weighted Results**	**Basic Demo-graphic Weights**	**Raking**	**Propensity Model Weights**	**GREG Weights**
Alcohol Risk							
Risk in harming themselves when have 5+ alcoholic drinks once or twice a week	Very Risky	59.7%	61.8%	62.1%	59.3%†	62.1%‡ (62.8%‡‡)	63.2%! (63.0%!!)
	Not Very Risky	40.3%	38.2%	37.9%	40.7%	37.9% (37.2%‡‡)	36.8% (37.0%)
Average age for first drink		17.2	18.0	18.2	17.7	18.0‡ 17.6‡‡	17.6! 17.5!!
Attitude towards College							
College is an investment in the future	Agree	97.2%	91.4%	88.5%	90.0%††	89.9%	88.5%
	Neither	2.0%	5.8	7.2%	4.7%	4.7%	6.2%
	Disagree	0.8%	2.8	4.4%	5.3%	5.5%	5.3%
A college degree is more important now than it used to be	Agree	86.6%	81.5	80.2%	82.3%	82.1%	79.9%
	Neither	5.5%	11.7	12.5%	8.3%	8.4%	11.8%
	Disagree	8.1%	6.9	7.3%	9.4%	9.5%	8.2%

† GSS Attitudinal, Cognitive and Conative questions along with HINTS Technology Adoption questions
†† GSS Attitudinal, Cognitive and Conative questions
‡ Propensity model – Demos raking adjustment
‡ ‡ Propensity model – GSS, Demos raking along with HINTS Technology Adjustment
! GREG GSS Attitudinal, Cognitive and Conative questions along with EMail Time
!! GREG GSS Attitudinal, Cognitive and Conative questions along with EMail Time & Technology Adoption questions

4. Discussion

This study is really a cautionary tale. We proposed the ABC model to provide a basis for identifying auxiliary measures to adjust survey results from non-probability samples. In the Alcohol Risk data study, the results look very promising. The adjustment when including general ABC questions from the GSS moved the results from the non-probability sample in the direction of the targets. However, questions from other studies should be equivalent to the Affective, Cognitive, or Cognitive questions were not as effective. The patterns and underlying structure of the data is still not understood, so additional research is necessary to understand why one set of questions is effective for one study but not another.

In the Attitudes towards College data, the methods are asked to correct for mode effect as well as for the effects from non-probability sampling. Recall that the study was supposed to determine whether or not one can transition an RDD telephone study to an online panel study. The data show significant mode effects. Even though the GSS questions showed similar mode response effects as the Attitudes toward College questions, the calibration processes did not overcome both the mode effect along with a possible nonprobability sample effect. There was some improvement. In one side analysis, an attitudinal variable central to the study was included in the adjustment. This adjustment “fixed” the response distributions of the most of the questions. This is an indication that the mode effect dominates any effect from the non-probability sample.

In addition to looking at the choice of questions, the study compared the efficacy of using a different adjustment method. The methods examined were Raking-Ratio, Propensity Score and GREG weighting. We observed improvement when using raking and propensity weighting. GREG weighting was less successful. In running through different combinations of variables, it seems that deeper adjustments using these variables may be more successful. One method multi-level regression post-stratification may be the next step.

We have to note that the online results are not “too different” from the gold standard results. Common across both studies was that adjustment by demographic variables alone did not reduce any bias, and if fact, worsened any bias. One observation is that the set of ABC measures used here did provide some adjustment for the attitudinal and behavioral measures in the two surveys. What is yet unclear is whether or not one needs different models for different measures within the same study. Public opinion polling already employs likely voter models along with political identification to adjust results for both online and telephone studies. These models are most often a combination of voter intention, past voting behavior and party identification. These can be viewed in the light of the ABC model. Party identification fulfills the affective or feelings component, past voting behavior fulfills the conative or action component, and voter intention can be assigned to cognitive or belief component. Future research needs to refine parameters for choosing questions and methods to incorporate survey specific attitude models.

References

AAPOR (2012) *AAPOR Statement: Understanding a “credibility interval” and how it differs from the “margin of sampling error” in a public opinion poll*.

Bagozzi, R. P., Tybout, A. M. , Craig, C. S. and Sternthal, B. (1979). “Construct Validity of the Tripartite Classification of Attitudes” *Journal of Marketing Research*, 16, 1, 88-95.

Baker, R., Brick, J.M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J.A., Gile, K. J., Tourangeau, R. (2013). “Summary Report of the AAPOR Task Force on Non-Probability Sampling.” *Journal of Survey Statistics and Methodology*, 1, 90-143.

Deming, W.E. (1943). *Statistical adjustment of data*. Wiley, New York.

Deming, W.E. and F. F. Stephan (1940). On a least squares adjustment of a sampled frequency when the expected marginals are known. *The Annals of Mathematical Statistics, 11, 427- 444*

Deville, J.-C. (2000) Generalized Calibration and Application to Weighting for Non-response. Proceedings of the 14th Compstat Symposium in Computational Statistics, Utrecht, Physica-Verlag.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382

Deville, Jean-Claude and Tille, Yves (2005). “Variance approximation under balanced sampling,” *Journal of Statistical Planning and Inference,* 128, 569 – 591.

Doe, P. and Smith, R. (2013) “Best of Both Worlds, can we make convenience samples representative?” ESOMAR 2013 Congress.

Gelman, A. and T.C. Little. 1997. Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* 23: 127{135).

Greenwald, A. (1968) "On defining attitude and attitude theory." In A. Greenwald, T. Brock, and T. Ostrom, eds., *Psychology Foundations of Attitudes. New York: Academic Press.*

Groeneman, S. (1994) "Multi-purpose household panels and general samples: How similar and how different." A*nnual meeting of the American Association for Public Opinion Research, Danvers, MA*.

Izrael, D., D.C. Hoaglin and M.P. Battaglia. (2000). A SAS macro for balancing a weighted sample. Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference, SAS Institute Inc., Cary, NC. pp. 1350–1355.

Kadane, J.B. and Seidenfeld, T. (1990) "Randomization in a Bayesian perspective" *Journal of Statistical Planning and Inference,* 25:329-345.

Kalton, G. (1983). *Compensating for Missing Survey Data*. Survey Research Center, Institute for Social Research, University of Michigan.

Putnam, R. D. (2000) *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon & Schuster

Rosenbaum, P.R. (2002). Observational Studies. New York: Spring-Verlag.

Rosenberg, M. and Hovland, C. (1960). "Cognitive, Affective and Behavioral Components of Attitudes." In Posenberg, et al. eds. *Attitude Organization and change*. New Haven, CT.: Yale University Press.

Roshwalb, A., El-Dash, N., and Young, C.A. (2012). "Towards the Use of Bayesian Credibility Intervals in Online Survey Result." Ipsos White Paper.

Rubin, D.B (2006). *Matched Sampling for Observational Studies*. Cambridge: Cambridge University press.

Sudman, S. and Wansink, B. (2002) Consumer *Panels*, South-Western Educational Publishing.

Thurstone, L. (1931). "The measurement of social attitudes." *Journal of Abnormal and Social psychology*, 19, 249-269.